Mellanox OFED for Linux Installation Guide

Rev 1.5.2-2.1.0

www.mellanox.com

NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT (“PRODUCT(S)”) AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES “AS-IS” WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCTO(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Mellanox Technologies
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403

Mellanox Technologies, Ltd.
PO Box 586 Hermon Building
Yokneam 20692
Israel
Tel: +972-4-909-7200
Fax: +972-4-959-3245

© Copyright 2011. Mellanox Technologies. All rights reserved.
Mellanox®, BridgeX®, ConnectX®, InfiniBlast®, InfiniBridge®, InfiniHost®, InfiniRISC®, InfiniScale®, InfiniPCI®, and PhyX® and Virtual Protocol Interconnect® are registered trademarks of Mellanox Technologies, Ltd.
CORE-Direct™ and FabricIT™ are trademarks of Mellanox Technologies, Ltd.

Table of Contents

Table of Contents 3
Chapter 1 Introduction 4
1.1 Hardware and Software Requirements 4
1.1.1 Hardware Requirements 4
Chapter 2 Downloading Mellanox OFED 5
Chapter 3 Installing Mellanox OFED 6
3.1 Pre-installation Notes 6
3.2 Installation Script 7
3.3 Installation Procedure 8
3.4 Installation Results 12
3.5 Post-installation Notes 14
Chapter 4 Updating Firmware After Installation 15
Chapter 5 Uninstalling Mellanox OFED 17

Mellanox Technologies Confidential

1 Introduction

This chapter describes how to install and test the Mellanox OFED for Linux package on a single host machine with Mellanox InfiniBand and/or Ethernet adapter hardware installed. The chapter includes the following sections:

- Section 1.1,"Hardware and Software Requirements," on page 4
- Section 2,"Downloading Mellanox OFED," on page 5
- Section 3,"Installing Mellanox OFED," on page 6
- Section 5,"Uninstalling Mellanox OFED," on page 17

1.1 Hardware and Software Requirements

1.1.1 Hardware Requirements

Platforms

- A server platform with an adapter card based on one of the following Mellanox Technologies' InfiniBand HCA devices:
  - MT25408 ConnectX®-2 (VPI, IB, EN, FCoE) (firmware: fw-ConnectX2)
  - MT25408 ConnectX® (VPI, IB, EN, FCoE) (firmware: fw-25408)
  - MT25208 InfiniHost® III Ex (firmware: fw-25218 for Mem-Free cards, and fw-25208 for cards with memory)
  - MT25204 InfiniHost® III Lx (firmware: fw-25204)
  - MT23108 InfiniHost® (firmware: fw-23108)

Note: For the list of supported architecture platforms, please refer to the *Mellanox OFED Release Notes* file.

Required Disk Space for Installation

- 400 MB

Software Requirements

Operating System

- Linux operating system

Note: For the list of supported operating system distributions and kernels, please refer to the *Mellanox OFED Release Notes* file.

Installer Privileges

- The installation requires administrator privileges on the target machine

Mellanox Technologies Confidential

2 Downloading Mellanox OFED

Step 1. Verify that the system has a Mellanox network adapter (HCA/NIC) installed by ensuring that you can see ConnectX or InfiniHost entries in the display.

The following example shows a system with an installed Mellanox HCA:

```
host1# lspci -v | grep Mellanox
02:00.0 InfiniBand: Mellanox Technologies MT25418 [ConnectX IB DDR, PCIe 2.0 2.5GT/s] (rev
a0)
```

Step 2. Download the ISO image to your host.

The image's name has the format MLNX_OFED_LINUX-<ver>-<OS label>.iso. You can download it from http://www.mellanox.com > Products > IB SW/Drivers.

Step 3. Use the md5sum utility to confirm the file integrity of your ISO image. Run the following command and compare the result to the value provided on the download page.

```
host1$ md5sum MLNX_OFED_LINUX-<ver>-<OS label>.iso
```


Mellanox Technologies Confidential

3 Installing Mellanox OFED

The installation script, `mlnxofedinstall`, performs the following:

- Discovers the currently installed kernel
- Uninstalls any software stacks that are part of the standard operating system distribution or another vendor's commercial stack
- Installs the MLNX_OFED_LINUX binary RPMs (if they are available for the current kernel)
- Identifies the currently installed InfiniBand and Ethernet network adapters and automatically[1] upgrades the firmware

3.1 Pre-installation Notes

- The installation script removes all previously installed Mellanox OFED packages and re-installs from scratch. You will be prompted to acknowledge the deletion of the old packages.

  Note: Pre-existing configuration files will be saved with the extension ".conf.saverpm".

- If you need to install Mellanox OFED on an entire (homogeneous) cluster, a common strategy is to mount the ISO image on one of the cluster nodes and then copy it to a shared file system such as NFS. To install on all the cluster nodes, use cluster-aware tools (such as pdsh).
- If your kernel version does not match with any of the offered pre-built RPMs, you can add your kernel version by using the "`mlnx_add_kernel_support.sh`" script located under the docs/ directory.

  Usage:

```
mlnx_add_kernel_support.sh -i|--iso <mlnx iso>
[-t|--tmpdir <local work dir>][-v|--verbose]
```

Example

The following command will create a MLNX_OFED_LINUX ISO image for RedHat 5.4 under the
/tmp directory.

```
MLNX_OFED_LINUX-1.5.1-rhel5.4/docs/mlnx_add_kernel_support.sh -i
/mnt/MLNX_OFED_LINUX-1.5.1-rhel5.4.iso
All Mellanox, OEM, OFED, or Distribution IB packages will be removed.
Do you want to continue?[y/N]:y
Removing OFED RPMs...
Running mkisofs...
Created /tmp/MLNX_OFED_LINUX-1.5.1-rhel5.4.iso
```

1. The firmware will not be updated if you run the install script with the '--without-fw-update' option.

Mellanox Technologies Confidential

3.2 Installation Script

Mellanox OFED includes an installation script called `mlnxofedinstall`. Its usage is described below. You will use it during the installation procedure described in Section 3.3, "Installation Procedure," on page 8.

Usage

```
./mlnxofedinstall [OPTIONS]
```

Note: If no options are provided to the script, then all available RPMs are installed.

Options

```
-c|--config <packages config_file>
                    Example of the configuration file can be found under docs
-n|--net <network config file>
                    Example of the network configuration file can be found
                    under docs
-p|--print-available Print available packages for the current platform and cre-
                    ate a corresponding ofed.conf file. The installation script
                    exits after creating ofed.conf.
--with-fc           Install FCoE support — Available on RHEL5.2 ONLY
--with-32bit        Install 32-bit libraries (default). This is relevant for
                    x86_64 and ppc64 platforms.
--without-32bit     Skip 32-bit libraries installation
--without-ib-bonding Skip ib-bonding RPM installation
--without-depcheck  Skip Distro's libraries check
--without-fw-update Skip firmware update
--force-fw-update   Force firmware update
--force             Force installation (without querying the user)
--all               Install all kernel modules, libibverbs, libibumad, librd-
                    macm, mft, mstflint, diagnostic tools, OpenSM, ib-bonding,
                    MVAPICH, Open MPI, MPI tests, MPI selector, perftest,
                    sdpnetstat and libsdp srptools, rds-tools, static and
                    dynamic libraries
--hpc               Install all kernel modules, libibverbs, libibumad, librd-
                    macm, mft, mstflint, diagnostic tools, OpenSM, ib-bonding,
                    MVAPICH, Open MPI, MPI tests, MPI selector, dynamic librar-
                    ies
--basic             Install all kernel modules, libibverbs, libibumad, mft,
                    mstflint, dynamic libraries
--msm               Install all kernel modules, libibverbs, libibumad, mft,
                    mstflint, diagnostic tools, OpenSM, ib-bonding, dynamic
                    libraries

                    NOTE: With --msm flag, the OpenSM daemon is configured to
                    run upon boot.
-v|-vv|-vvv         Set verbosity level
-q                  Set quiet - no messages will be printed
```


Mellanox Technologies Confidential

mlnxofedinstall Return Codes

Table 1 lists the `mlnxofedinstall` script return codes and their meanings.

Table 1 - `mlnxofedinstall` ***Return Codes***

Return Code	Meaning
0	The Installation ended successfully
1	The installation failed
2	No firmware was found for the adapter device
3	Failed to start the `mst` driver

3.3 Installation Procedure

Step 1. Login to the installation machine as root.

Step 2. Mount the ISO image on your machine

```
host1# mount -o ro,loop MLNX_OFED_LINUX-<ver>-<OS label>.iso /mnt
```

Note: After mounting the ISO image, `/mnt` will be a Read Only folder.

Step 3. Run the installation script

```
host1# /mnt/mlnxofedinstall
This program will install the MLNX_OFED_LINUX package on your machine.
Note that all other Mellanox, OEM, OFED, or Distribution IB packages will be removed.
Do you want to continue?[y/N]:y

Uninstalling the previous version of OFED

Starting MLNX_OFED_LINUX-1.5.2-0.0.5 installation ...

Installing kernel-ib RPM
Preparing...                ##################################################
kernel-ib                   ##################################################
Installing kernel-ib-devel RPM
Preparing...                ##################################################
kernel-ib-devel             ##################################################
Installing kernel-mft RPM
Preparing...                ##################################################
kernel-mft                  ##################################################
Installing mpi-selector RPM
Preparing...                ##################################################
mpi-selector                ##################################################
```


Mellanox Technologies Confidential

```
Install user level RPMs:
Preparing...                ###########################################
libibverbs                  ###########################################
libibumad                   ###########################################
libibverbs                  ###########################################
libibumad                   ###########################################
libibumad-devel             ###########################################
librdmacm                   ###########################################
opensm-libs                 ###########################################
librdmacm                   ###########################################
libibmad                    ###########################################
mvapich_intel               ###########################################
libibverbs-devel            ###########################################
openmpi_intel               ###########################################
libmverbs                   ###########################################
opensm-libs                 ###########################################
libmthca                    ###########################################
libmlx4                     ###########################################
libibcm                     ###########################################
libibmad                    ###########################################
libsdp                      ###########################################
dapl                        ###########################################
dapl                        ###########################################
ibutils2                    ###########################################
libmverbs                   ###########################################
ofed-scripts                ###########################################
libibverbs-devel            ###########################################
libibverbs-devel-static     ###########################################
libibverbs-devel-static     ###########################################
libibverbs-utils            ###########################################
libmthca                    ###########################################
libmthca-devel-static       ###########################################
libmthca-devel-static       ###########################################
libmlx4                     ###########################################
libmlx4-devel               ###########################################
libmlx4-devel               ###########################################
libmverbs-devel             ###########################################
libmverbs-devel             ###########################################
libmqe                      ###########################################
libmqe                      ###########################################
libibcm                     ###########################################
libibcm-devel               ###########################################
libibcm-devel               ###########################################
libibumad-devel             ###########################################
libibumad-static            ###########################################
libibumad-static            ###########################################
libibmad-devel              ###########################################
libibmad-devel              ###########################################
libibmad-static             ###########################################
libibmad-static             ###########################################
```


Mellanox Technologies Confidential

```
ibsim                     ##################################################
librdmacm-utils           ##################################################
librdmacm-devel           ##################################################
librdmacm-devel           ##################################################
libsdp                    ##################################################
libsdp-devel              ##################################################
libsdp-devel              ##################################################
opensm                    ##################################################
opensm-devel              ##################################################
opensm-devel              ##################################################
opensm-static             ##################################################
opensm-static             ##################################################
compat-dapl               ##################################################
compat-dapl               ##################################################
compat-dapl-devel         ##################################################
compat-dapl-devel         ##################################################
dapl-devel                ##################################################
dapl-devel                ##################################################
dapl-devel-static         ##################################################
dapl-devel-static         ##################################################
dapl-utils                ##################################################
perftest                  ##################################################
mstflint                  ##################################################
mft                       ##################################################
sdpnetstat                ##################################################
srptools                  ##################################################
rds-tools                 ##################################################
ibutils                   ##################################################
cc_mgr                    ##################################################
ibdump                    ##################################################
infiniband-diags          ##################################################
qperf                     ##################################################
mlnxofed-docs             ##################################################
mvapich_gcc               ##################################################
mvapich_pgi               ##################################################
openmpi_gcc               ##################################################
openmpi_pgi               ##################################################
mpitests_mvapich_gcc      ##################################################
mpitests_mvapich_pgi      ##################################################
mpitests_mvapich_intel    ##################################################
mpitests_openmpi_gcc      ##################################################
mpitests_openmpi_pgi      ##################################################
mpitests_openmpi_intel    ##################################################
Device (15b3:634a):
        02:00.0 InfiniBand: Mellanox Technologies MT25418 [ConnectX VPI PCIe 2.0 2.5GT/s -
IB DDR / 10GigE] (rev a0)
         Link Width: 8x
         Link Speed: 2.5Gb/s
```

```
Installation finished successfully.

Programming HCA firmware for /dev/mst/mt25418_pci_cr0 device
Running: mlxburn -d /dev/mst/mt25418_pci_cr0 -fw /mnt/firmware/fw-25408/2_8_0000/fw-25408-rel.mlx -dev_type 25408 -no
-I- Querying device ...
-I- Using auto detected configuration file: /mnt/firmware/fw-25408/2_8_0000/MHGH28-XTC_A4-A7.ini (PSID = MT_04A0140005)
-I- Generating image ...
Current FW version on flash: 2.7.0
New FW version: 2.8.0
Burning FW image without signatures - OK Restoring signature - OK
-I- Image burn completed successfully.
Please reboot your system for the changes to take effect.
warning: /etc/infiniband/openib.conf saved as /etc/infiniband/openib.conf.rpmsave
```

Note In case your machine has the latest firmware, no firmware update will occur and the installation script will print at the end of installation a message similar to the following:

```
...
Installation finished successfully.
The firmware version 2.8.0000 is up to date.
Note: To force firmware update use '--force-fw-update' flag.
```

Note In case your machine has an unsupported network adapter device, no firmware update will occur and the error message below will be printed. Please contact your hardware vendor for help on firmware updates.

Error message:

```
-I- Querying device ...
-E- Can't auto detect fw configuration file: ...
```

Step 4. In case the installation script performed firmware updates to your network adapter hardware, it will ask you to reboot your machine.

Step 5. The script adds the following lines to `/etc/security/limits.conf` for the userspace components such as MPI:

```
* soft memlock unlimited
* hard memlock unlimited
```

These settings unlimit the amount of memory that can be pinned by a user space application.
If desired, tune the value unlimited to a specific amount of RAM.

Step 6. For your machine to be part of the InfiniBand/VPI fabric, a Subnet Manager must be running on one of the fabric nodes. At this point, Mellanox OFED for Linux has already installed the OpenSM Subnet Manager on your machine.

Mellanox Technologies Confidential

Step 7. (InfiniBand only) Run the `hca_self_test.ofed` utility to verify whether or not the InfiniBand link is up. The utility also checks for and displays additional information such as

- HCA firmware version
- Kernel architecture
- Driver version
- Number of active HCA ports along with their states
- Node GUID

Note For more details on `hca_self_test.ofed`, see the file `hca_self_test.readme` under `docs/`.

```
host1# /usr/bin/hca_self_test.ofed
---- Performing InfiniBand HCA Self Test ----
Number of HCAs Detected ..............  1
PCI Device Check .....................  PASS
Kernel Arch ..........................  x86_64
Host Driver Version ..................  MLNX_OFED_LINUX-1.5.2-0.0.5 (OFED-1.5.2-20101014-
1355): 1.5.2-2.6.32.12_0.7_default
Host Driver RPM Check ................  PASS
HCA Firmware on HCA #0 ...............  v2.8.0000
HCA Firmware Check on HCA #0 ..........  PASS
Host Driver Initialization ............  PASS
Number of HCA Ports Active ............  1
Port State of Port #1 on HCA #0 ........  UP 4X DDR
Port State of Port #2 on HCA #0 ........  INIT
Error Counter Check on HCA #0 ..........  PASS
Kernel Syslog Check ...................  PASS
Node GUID on HCA #0 ...................  00:02:c9:03:00:00:10:e0
------------------ DONE --------------------
```

Note After the installer completes, information about the Mellanox OFED installation such as prefix, kernel version, and installation parameters can be retrieved by running the command `/etc/infiniband/info`.

3.4 Installation Results

Software

- The OFED and MFT packages are installed under the /usr directory.
- The kernel modules are installed under:
  - InfiniBand subsystem:

    ```
    /lib/modules/`uname -r`/updates/kernel/drivers/infiniband/
    ```

  - mlx4 driver:

    ```
    Under /lib/modules/`uname -r`/updates/kernel/drivers/net/mlx4
     you will find mlx4_core.ko,  mlx4_en.ko,  mlx4_ib.ko (and mlx4_fc if
     you ran the installation script with --with-fc)
    ```


Mellanox Technologies Confidential

- RDS:

  `/lib/modules/`uname -r`/updates/kernel/net/rds/rds.ko`

- Bonding module:

  `/lib/modules/`uname -r`/updates/kernel/drivers/net/bonding/bonding.ko`

- The package kernel-ib-devel include files are placed under `/usr/src/ofa_kernel/include/`. These include files should be used when building kernel modules that use the stack. (Note that the include files, if needed, are "backported" to your kernel.)
- The raw package (un-backported) source files are placed under `/usr/src/ofa_kernel-<ver>`
- The script `openibd` is installed under `/etc/init.d/`. This script can be used to load and unload the software stack.
- The script `connectx_port_config` is installed under `/sbin`. This script can be used to configure the ports of ConnectX network adapter cards to Ethernet and/or InfiniBand.
- The directory `/etc/infiniband` is created with the files `info` and `openib.conf and connectx.conf`. The `info` script can be used to retrieve Mellanox OFED installation information. The `openib.conf` file contains the list of modules that are loaded when the `openibd` script is used. The `connectx.conf` file saves the ConnectX adapter card's ports configuration to Ethernet and/or InfiniBand. This file is used at driver start/restart (`/etc/init.d/openibd start`) .
- The file `90-ib.rules` is installed under `/etc/udev/rules.d/`
- If OpenSM is installed, the daemon `opensmd` is installed under `/etc/init.d/` and `opensm.conf` is installed under `/etc`.
- If IPoIB configuration files are included, `ifcfg-ib<n>` files will be installed under:

  - `/etc/sysconfig/network-scripts/` on a RedHat machine
  - `/etc/sysconfig/network/` on a SuSE machine

- The installation process unlimits the amount of memory that can be pinned by a user space application. See Step 5.
- Man pages will be installed under `/usr/share/man/`

Firmware

- The firmware of existing network adapter devices will be updated if the following two conditions are fullfilled:

  1. You run the installation script in default mode; that is, *without* the option '--without-fw-update'.
  2. The firmware version of the adapter device is older than the firmware version included with the Mellanox OFED ISO image

  Note: If an adapter's Flash was originially programmed with an Expansion ROM image, the automatic firmware update will also burn an Expansion ROM image.

- In case your machine has an unsupported network adapter device, no firmware update will occur and the error message below will be printed. Please contact your hardware vendor for help on firmware updates.

  Error message:

Mellanox Technologies Confidential

```
-I- Querying device ...
-E- Can't auto detect fw configuration file: ...
```

3.5 Post-installation Notes

- Most of the Mellanox OFED components can be configured or reconfigured after the installation by modifying the relevant configuration files. See the relevant chapters in this manual for details.
- The list of the modules that will be loaded automatically upon boot can be found in the `/etc/infiniband/openib.conf` file.

Mellanox Technologies Confidential

4 Updating Firmware After Installation

In case you ran the `mlnxofedinstall` script with the '`--without-fw-update`' option and now you wish to (manually) update firmware on you adapter card(s), you need to perform the following steps:

Note: The following steps are also appropriate in case you wish to burn newer firmware that you have downloaded from Mellanox Technologies' Web site (http://www.mellanox.com > Downloads > Firmware).

Step 1. Start mst.

```
host1# mst start
```

Step 2. Identify your target InfiniBand device for firmware update.

a. Get the list of InfiniBand device names on your machine.

```
host1# mst status
MST modules:
------------
    MST PCI module loaded
    MST PCI configuration module loaded
    MST Calibre (I2C) module is not loaded

MST devices:
------------
/dev/mst/mt25418_pciconf0        - PCI configuration cycles access.
                                   bus:dev.fn=02:00.0 addr.reg=88 data.reg=92
                                   Chip revision is: A0
/dev/mst/mt25418_pci_cr0         - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdef00000 size=0x100000
                                   Chip revision is: A0
/dev/mst/mt25418_pci_msix0       - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdeefe000 size=0x2000
/dev/mst/mt25418_pci_uar0        - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdc800000 size=0x800000
```

b. Your InfiniBand device is the one with the postfix "_pci_cr0". In the example listed above, this will be `/dev/mst/mt25418_pci_cr0`.

Step 3. Burn firmware.

a. Burning a firmware binary image using **`mstflint`** (that is already installed on your machine).

Please refer to `MSTFLINT_README.txt` under `docs/`.

b. Burning a firmware image from a .mlx file using the **`mlxburn`** utility (that is already installed on your machine).

The following command burns firmware onto the ConnectX device with the device name obtained in the example of Step 2.

```
host1$ mlxburn -dev /dev/mst/mt25418_pci_cr0 \
-fw /mnt/firmware/fw-25408/fw-25408-rel.mlx
```


Mellanox Technologies Confidential

Warning! Make sure that you have the correct *device name*, *firmware path*, and *firmware file name* before running this command. For help, please refer to the *Mellanox Firmware Tools (MFT) User's Manual* under `/mnt/docs/`.

Step 3. Reboot your machine after the firmware burning is completed.

Mellanox Technologies Confidential

5 Uninstalling Mellanox OFED

Use the script `/usr/sbin/ofed_uninstall.sh` to uninstall the Mellanox OFED package. The script is part of the `ofed-scripts` RPM.

Mellanox Technologies Confidential